Physio-DPO：将大语言模型与蛋白质能量景观对齐，以消除结构幻觉

Qiwei Meng西安交通大学
Twilight_M@stu.xjtu.edu.cn

arXiv:2601.00647v1 [cs.CL] 2 Jan 2026

摘要

大型蛋白质语言模型在生成式蛋白质设计方面展现出巨大潜力，但仍频繁产生结构幻觉，生成语言似然度高、却会折叠成热力学不稳定构象的序列。现有的对齐方法（如直接偏好优化）在此场景下存在局限，因为它们将偏好建模为二元标签，并忽略了物理能量景观的连续结构。我们提出Physio-DPO，这是一种基于物理信息的对齐框架，可使蛋白质语言模型以热力学稳定性为基础。Physio-DPO 引入了一种幅度感知目标，根据天然结构与物理扰动困难负样本之间的能量差来调整优化更新幅度。实验表明，Physio-DPO 的表现始终优于包括监督微调、近端策略优化和标准 DPO 在内的强基线，将自一致性 RMSD 降至 1.28 埃，并将可折叠性提高至 92.8%。定性分析进一步表明，Physio-DPO 能够通过恢复疏水核心堆积和氢键网络等生物物理相互作用，有效缓解结构幻觉。

1 引言

蛋白质语言模型（蛋白质语言模型）规模化方面的最新进展，以 ESM系列（Lin 等人，2023；Hayes 等人，2024年）和 ProGen（Madani 等人，2023）为代表，显著推动了计算蛋白质设计的发展。这些模型通过从数十亿条序列中内化进化的统计语法，展现出强大的生成能力，能够从头生成类蛋白质序列。然而，一项根本性的错位仍然存在。蛋白质语言模型的训练目标是最小化词元级困惑度，这仅是进化适应度的间接替代指标，并未明确优化热力学稳定性。因此，即便是大规模预训练语言模型，也经常生成结构幻觉：模型对这些序列表现出高度信心，但它们却折叠成高能量、物理上无效的构象，其特征包括无序区域、空间位阻冲突或暴露的疏水核心（Anishchenko 等人，2021年；Gopalan 和 Narayanan，2025年）。

直接偏好优化（DPO）（Rafailov 等人，2023）近年来逐渐成为一种更稳定的离线替代方案，它将强化学习重新表述为针对偏好对的分类目标。尽管 DPO 已被证明在对齐方面行之有效（Das 等人，2025年），但其标准形式并不适用于生物物理优化。DPO 将偏好建模为二元关系，舍弃了候选结构之间质量差异的幅度。然而，在物理系统中，天然结构与诱饵结构之间的能量差承载着有关能量景观拓扑和陡峭程度的重要信息。将连续的热力学信号简化为二元标签，会使模型无法区分细微波动与严重的结构失效。此外，SimPO（Meng 等人，2024年）和 ORPO（Hong 等人，2024年）等无需参考模型的变体取消了显式正则化，因此可能侵蚀支撑生物学合理性的进化先验。

为解决离散偏好学习与连续物理定律之间的不匹配问题，我们提出**Physio-DPO**，这是一种物理信息对齐框架，旨在使大规模预训练语言模型扎根于热力学现实。Physio-DPO 在标准 DPO 的基础上扩展出一种感知幅度的目标函数，根据物理能量差明确调整优化更新的权重，使模型能够集中能力解决显著的稳定性障碍。我们还提出了一种困难负样本挖掘策略，用于生成具有语言上的合理性、但结构上不稳定的对抗性诱饵结构，迫使模型学习细粒度的生物物理学差异。针对蛋白质生成的大量实验表明

物理信息偏好优化能够在前所未有的模型规模下，实现稳定、可扩展且以物理规律为基础的蛋白质生成。

我们的贡献总结如下：**(1)** 我们构建了一个大规模、以物理为基础的偏好数据集，包含 100 万组天然态 – 诱饵态配对，其中的困难负样本通过有针对性的物理扰动生成，以揭示细微却关键的结构失效。**(2)** 我们提出了一种物理信息偏好优化框架，在标准 DPO 的基础上扩展出连续的幅度感知目标函数，使梯度更新幅度能够随热力学能隙变化。**(3)** 我们提供了理论分析，表明所提出的能量加权目标能够降低梯度方差，并且等价于优化底层物理能量分布的一种有原则的代理目标。**(4)** 在大规模蛋白质生成任务上的广泛实验表明，Physio-DPO 始终优于强基线，达到了最先进的结构准确性（sc-RMSD 为 **1.28 埃**），同时大幅减轻了结构幻觉。

2 相关工作

蛋白质生成模型与幻觉。 早期的蛋白质生成方法依赖从多序列比对（MSA）中提取的统计相关性，例如 Potts模型（Levy 等，2017年）。随着 Transformer架构（Vaswani 等，2017年）的规模不断扩大，研究范式转向了在海量宏基因组数据库上训练的自回归蛋白质语言模型（Rives 等，2021年；Elnaggar 等，2021年）。ESM 系列（Hayes 等，2024年；Lin 等，2023）和 ProtGPT2（Ferruz 等，2022）等模型能够捕捉远程进化依赖关系，从而生成多样化的序列。然而，这些模型优化的是词元级交叉熵损失，它虽可作为进化适应度的替代指标，却不能直接衡量结构稳定性。因此，这些模型容易产生幻觉：生成的序列在统计上看似合理，却会因空间位阻冲突或未满足的氢键而无法折叠成明确的三级结构（Anishchenko 等，2021年）。扩散模型（Watson 等，2023；Ingraham 等，2023；Lisanza 等，2025）虽然能够显式生成结构，但计算成本高昂，且缺乏蛋白质语言模型所具备的序列设计灵活性。我们的工作保留了蛋白质语言模型的高效性，同时通过对齐来确保结构有效性。

对齐。 传统上，将语言模型对齐至特定目标主要依赖基于人类反馈的强化学习（Zhou 等人，2025年；Mei 等人，2025年），最常见的实现方式是近端策略优化（PPO）（Schulman 等人，2017年）。在蛋白质设计中，强化学习已用于优化溶解度或荧光等特性（Angermueller 等人，2019年）；然而，近端策略优化需要分别训练奖励网络和价值网络，往往导致训练不稳定且内存开销巨大。直接偏好优化（DPO）（Rafailov 等人，2023）通过将强化学习重新表述为基于偏好的分类，提供了一种更稳定的离线替代方案。此后，面向自然语言处理的扩展方法，包括 IPO（Azar 等人，2024年）和 KTO（Ethayarajh 等人，2024年），改进了边际处理方式；而 SimPO（Meng 等人，2024年）和 ORPO（Hong 等人，2024年）等无需参考模型的方法则提高了效率。尽管如此，这些方法在生物序列设计方面仍不够理想：无需参考模型的方法会舍弃 KL正则化，因而有丧失进化先验、造成灾难性遗忘的风险；而且它们都将偏好视为二元标签，忽略了物理信号的大小。**Physio-DPO** 通过保留 KL锚定以维持生物学合理性，同时将热力学量值直接纳入偏好优化，克服了这些局限，从而弥合语义对齐与物理有效性之间的差距。

3 预备知识

直接偏好优化（DPO）。 直接偏好优化（DPO）（Rafailov 等人，2023）是一种无需强化学习、利用偏好数据对齐语言模型的框架。给定参考策略 π_{ref} 和可训练策略 π_θ，它通过最大化以下概率来优化 π_θ 优先选择赢家 y_w而非输家 y_l：

$$P(y_w \succ y_l) = \sigma(r(x, y_w) - r(x, y_l)), \quad (1)$$

其中，σ 是 Sigmoid函数，$r(x,\ y)$遵循 Bradley–Terry 模型，从而得到一个稳定的、基于偏好的目标函数，并由 π_{ref}进行正则化。

问题形式化。 我们将蛋白质设计形式化为一个无条件序列生成问题。令 x 表示一个通用前缀（例如，起始标记）；由 θ参数化的自回归预训练语言模型定义了一个在氨基酸序列$y \in \mathcal{Y}$上的策略 $\pi_\theta(y|x)$。尽管预训练蛋白质语言模型能够捕捉进化上的合理性，但它们并未明确确保生物物理有效性。我们假设可以访问一个物理

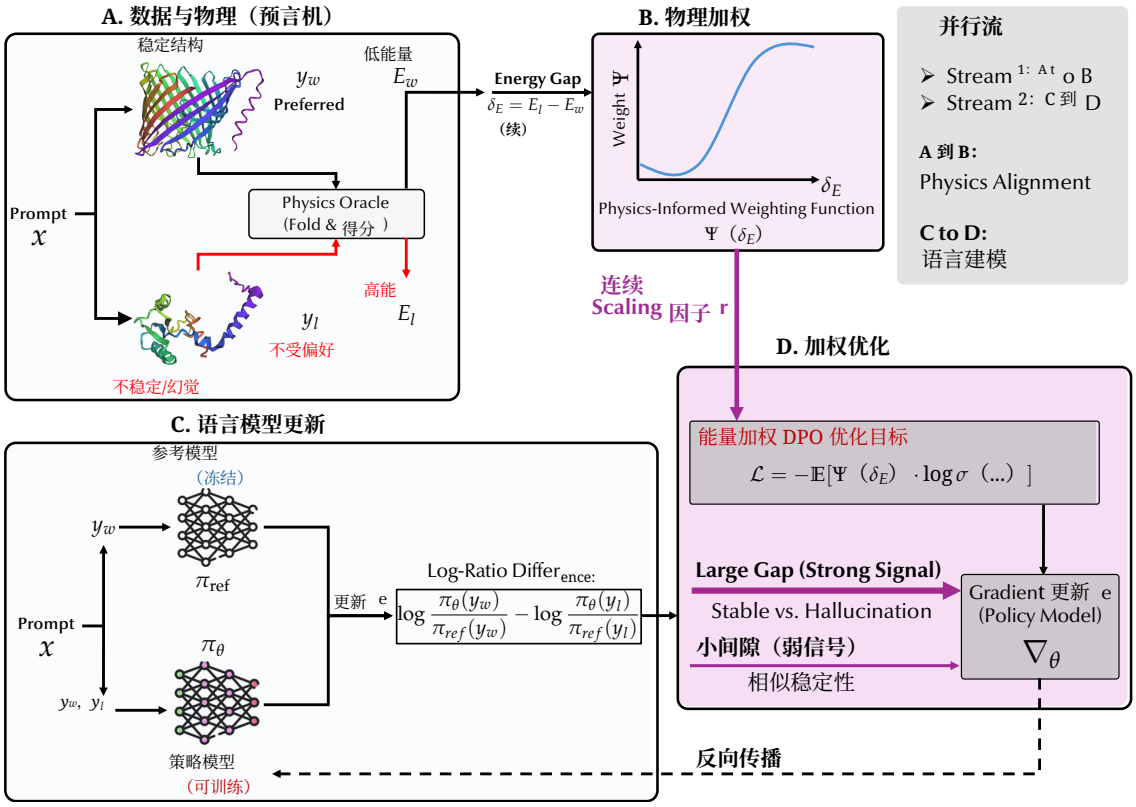

图 1：Physio-DPO 框架概览。物理流折叠一个采样对（y_w，y_l），并计算能量差 δ_E，再将其映射为权重 $\Psi(\delta_E)$。语言建模流使用策略 π_θ 和冻结参考模型 π_{ref}计算 DPO 对数比。Physio-DPO 根据 $\Psi(\delta_E)$ 对每个 DPO 项重新加权，放大来自稳定性差距较大的样本对的更新，并使模型与连续能量景观保持一致。

能量预言机 $\mathcal{E}(y)$，其中能量越低，热力学稳定性越高。我们的目标是调整π_θ，使其倾向于低能量序列，同时与参考模型 π_{ref} 保持接近，以维持多样性。

4 方法

虽然标准 DPO 提供了稳定的对齐目标，但它将偏好视为二元值，忽略了大小信息（Liu 等人，2025年）。在生物物理场景中，能量差反映结构不稳定性，这种离散化会舍弃连续能量景观中的关键信息。因此，我们提出 **Physio-DPO**，一种物理信息对齐框架，将热力学能量大小纳入偏好优化。如图1所示，Physio-DPO 包含两个阶段：（i）通过 Generate–Fold–Score 流程构建基于物理原理的偏好数据集（第4.1节）；以及（ii）优化能量加权目标，根据物理稳定性差距缩放 DPO 梯度（第4.2节）。

4.1 PhysioPref-1M 基准

蛋白质设计中的有效偏好对齐需要密集且以物理为依据的监督信号，而现有的指令微调数据集在很大程度上缺乏这类监督信号。为此，我们推出了 PhysioPref-1M，这是一个大规模偏好基准，包含依据热力学标准标注的 100 万个蛋白质对。该数据集通过对抗式生成与筛选流程（图 2）构建，该流程有意诱导并识别困难负样本——这些结构看似能够折叠，却违反了物理稳定性。通过将稳定蛋白质与不稳定或病态诱饵进行对比，构建偏好对，从而确保能量差具有信息量。进一步通过人工参与的评估验证了自动标注的可靠性。

4.2 优化目标

令 $\mathcal{E}(y)$ 表示物理能量。我们假设偏好强度是连续的，而非二元的，并由能量差上的玻尔兹曼分布决定。对于 $(y_w,\ y_l)$，偏好强度由能量差决定：

$$\delta_E(y_w, y_l) = \text{ReLU}\big(\mathcal{E}(y_l) - \mathcal{E}(y_w)\big) \qquad (2)$$

标准 DPO 最大化相对于参考模型的偏好响应对数似然。我们通过引入幅度感知公式扩展这一目标

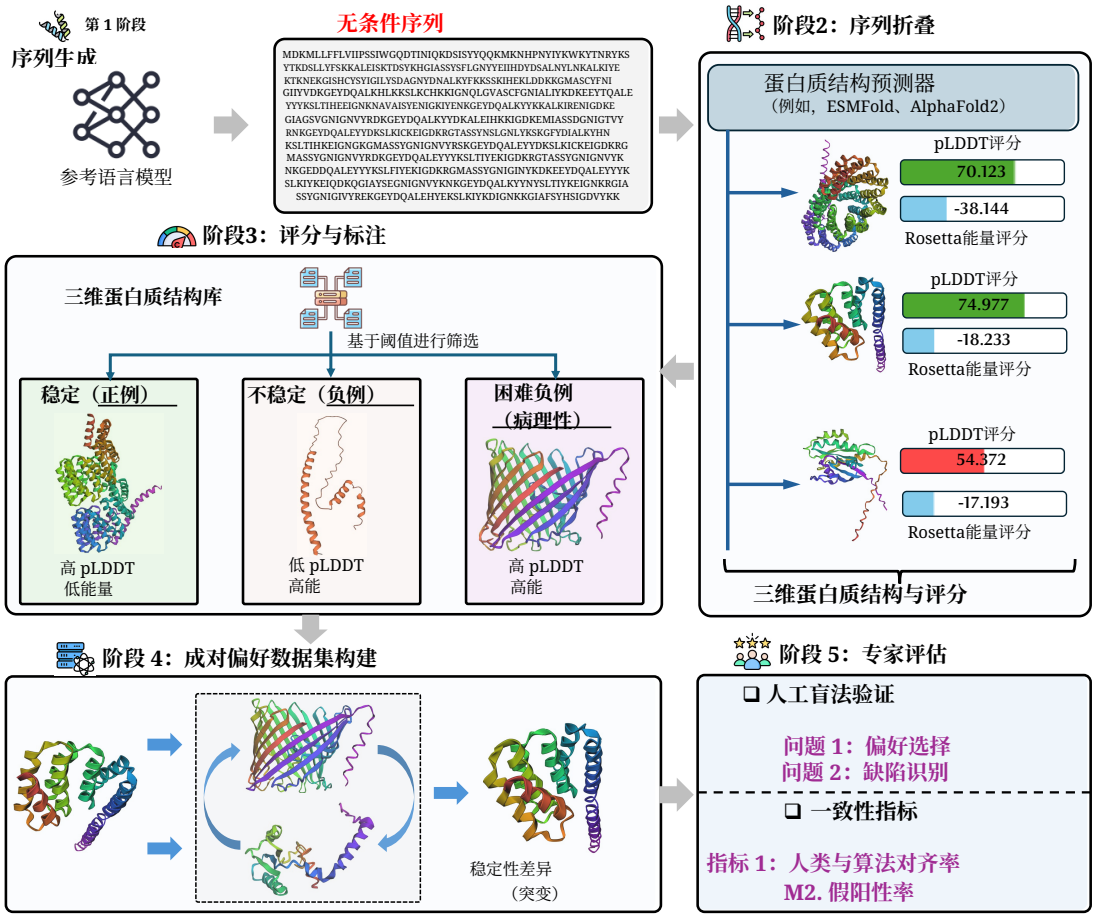

图2：PhysioPref-1M 的构建流程。**步骤1：**从参考语言模型中生成多样化序列（Ferruz 等人，2022）；**步骤2：**通过折叠进行结构预测；**步骤 3：**基于 pLDDT 置信度（ Fang 等人，2025 年）和 Rosetta 能量评分（ Alford 等人，2017年）进行评分和标注，包括识别置信度高但稳定性差的困难负例；**步骤 4：**构建能够最大化稳定性差距的偏好对；以及**步骤 5：**通过人在回路评估，验证标签与生物物理学判断的一致性。

，引入物理信息加权函数 $\Psi:\ \mathbb{R}^+ \to [0, \lambda_{\max}]$，将能量差映射为优化强度。

$$\Psi(\delta_E) = \lambda \cdot \sigma\left(\frac{\delta_E - \mu}{\tau}\right) \tag{3}$$

其中，μ 设定临界能量边界附近的灵敏度，τ 控制过渡的陡峭程度，抑制来自较小 δ_E 的噪声，同时放大来自困难负例的信号。

能量加权目标。 我们将$\Psi(\delta_E)$整合到直接偏好优化公式中。Physio-DPO目标函数定义如下：

$$\mathcal{L}_{\text{Physio}}(\pi_\theta; \pi_{\text{ref}}) = -\mathbb{E}_{(x,y_w,y_l)\sim\mathcal{D}}\left[\Psi(\delta_E) \cdot \log\sigma\left(\beta\log\frac{\pi_\theta(y_w|x)}{\pi_{\text{ref}}(y_w|x)} - \beta\log\frac{\pi_\theta(y_l|x)}{\pi_{\text{ref}}(y_l|x)}\right)\right]. \tag{4}$$

这可以解释为一种成本敏感学习方法，其中误分类成本是动态的，并由物理定律决定。

4.3 梯度调制分析

为了分析 Physio-DPO 如何提升稳定性，我们考察其梯度动态。令 $r_\theta(x,\ y) = \beta\log\frac{\pi_\theta(y|x)}{\pi_{\text{ref}}(y|x)}$ 表示隐式奖励。Physio-DPO 目标函数的梯度为：

$$\nabla_\theta\mathcal{L}_{\text{Physio}} = -\mathbb{E}[\Psi(\delta_E)\,\sigma(-\Delta r_\theta)\,\nabla_\theta\Delta r_\theta], \tag{5}$$

其中 $\Delta r_\theta = r_\theta(x,\ y_w) - r_\theta(x,\ y_l)$。这一分解揭示了一种梯度调制机制：一旦偏好被明确学会，标准 DPO误差项 $\sigma(-\Delta r_\theta)$便会消失，而物理信息增益 $\Psi(\delta_E)$则根据能量差调整更新幅度。因此，对于稳定性差异可忽略不计的模糊样本对，梯度会受到抑制；对于存在严重物理违规的困难负例，梯度则会被放大，从而使优化聚焦于最关键的生物物理错误。

理论洞见。我们通过理论分析表明，Physio-DPO会形成一种物理信息优化课程。在训练初期，Physio-DPO的梯度在局部上等价于最大化一个与物理能量差成正比的奖励函数，使隐式奖励与负能量景观保持一致。更一般地，可以证明，能量依赖加权项会放大存在严重物理违规的困难负样本对的梯度更新，同时抑制针对物理上模糊情形的更新。最后，我们建立了Physio-DPO与热力学平衡之间的联系，表明该目标近似于向由物理能量定义的玻尔兹曼分布进行KL最小化。

5 主要结果

我们在蛋白质生成任务上评估 Physio-DPO，以回答三个问题：（**Q1**）融入物理能量景观是否优于二元偏好对齐；（**Q2**）Physio-DPO 是否能够缓解结构幻觉，即高置信度但低稳定性的情况；以及（**Q3**）能量加权目标 $\Psi(\delta_E)$ 和困难负样本挖掘如何促进性能提升。

5.1 实验设置

数据集。我们使用自行构建的 PhysioPref-1M 基准（第 3.2 节）。我们依据序列一致性（使用 MMseqs2）严格划分数据集，将其分为训练集（900,000）、验证集（50,000）和测试集（50,000），确保没有测试序列与训练样本共享 $> 30\%$ 的一致性，从而评估泛化能力而非记忆化。

基线。我们将 Physio-DPO 与涵盖不同规模和对齐范式的一组基线进行比较，包括（**1**）**未对齐的蛋白质语言模型**：ProGen2-XL（6.4B）、ESM-3 Open（1.4B）和 ProtGPT2（762M）；（**2**）**监督微调（SFT）**：在稳定（y_w）子集上进行监督微调的 ProGen2-XL；（**3**）**基于强化学习的优化（PPO）**：我们使用 PPO 实施标准的基于人类反馈的强化学习，其中奖励模型在 PhysioPref-1M 偏好对上进行训练；（**4**）**偏好优化方法**：二元 DPO、IPO 和 KTO。

实现细节。所有模型均从 ProGen2-XL 初始化（Nijkamp 等人，2023）。我们使用 LoRA（Hu 等人，2022），秩为 $r = 16$，并使用 $\alpha = 32$ 进行微调。实验在 $4 \times$ NVIDIA A100（80GB）GPU 上，使用 HuggingFace TRL。我们将直接偏好优化系数设为$\beta = 0.1$。我们采用物理信息加权 Ψ（δ_E），缩放参数为 $\mu = 50$ 和 $\tau = 10$。

评估指标。我们从四个维度评估生成的蛋白质：**(1) 结构稳定性**，以自一致性RMSD（sc-RMSD）衡量；**(2) 可折叠性**，定义为预测 pLDDT 大于 70 的序列比例；**(3) 生物物理有效性**，以每个残基的平均 Rosetta能量量化；以及**(4) 多样性**，通过语言模型困惑度和与训练集的最大序列一致性进行评估。

5.2 结构与生物物理对齐

表 1报告了生成 3 万种新型蛋白质的结果。在未对齐的骨干模型中，ProGen2-XL 的基线表现最佳，可折叠性为 52.4%；但仍有近一半序列无法折叠，表明结构幻觉依然存在。对齐显著提升了生成质量：监督微调将可折叠性提高至 71.5%，而基于偏好的方法（DPO、IPO、KTO）则进一步提升至 80% 以上。尽管近端策略优化取得了有竞争力的结构指标，但其表现明显不稳定，困惑度偏高便反映了这一点。相比之下，Physio-DPO 的整体表现最佳：sc-RMSD 降低 0.54 埃，可折叠性提升至 92.8%，平均能量达到最低值（-3.05 相对能量单位），同时保留了语言多样性。这些结果表明，显式优化连续能量景观能够有效抑制二元偏好目标所忽略的困难负样本。

5.3 训练动态分析

为评估优化效率和稳定性，我们分析了训练期间的训练动态，跟踪物理能量、KL 散度和 sc-RMSD。如图3(a,b) 所示，近端策略优化在约 2K 步后变得不稳定，KL 散度迅速上升，能量表现恶化，表明策略因奖励利用而发生漂移。标准 DPO 收敛迅速，但很早便趋于饱和，限制了进一步改进。相比之下，Physio-DPO 在整个训练过程中持续改进，在保持 KL 散度有界（约 1.5 纳特）的同时，实现了最低的能量和 sc-RMSD。这表明能量加权目标能够有效正则化优化，并保持与预训练主干网络的对齐。

方法	结构指标 sc-RMSD（埃）↓	pLDDT ↑	可折叠性（%）↑	生物物理 能量（相对能量单位）↓	多样性指标 PPL ↓	序列一致性 (%) ↓
预训练骨干网络（零样本生成）						
ProtGPT2 (762M) (Ferruz 等人，2022)	5.12	48.5	22.1	-0.85	8.2	-
ESM-2 (3B)† (Lin 等人，2023)	4.55	56.2	35.4	-1.05	6.8	-
ESM-3 (1.4B)(Hayes 等人，2024)	3.95	62.5	45.8	-1.31	6.5	-
ProGen2-XL（6.4B）（Nijkamp 等人，2023）	3.25	67.8	52.4	-1.65	**6.1**	-
对齐方法（骨干模型：*ProGen2-XL 6.4B + LoRA*）						
监督微调（监督微调）	2.35	75.2	71.5	-2.25	7.4	34.1
近端策略优化（基于人类反馈的强化学习）（Schulman 等人，2017年）	2.15	78.5	79.2	-2.48	12.5	39.8
直接偏好优化（标准）（Rafailov 等人，2023）	1.82	81.3	83.6	-2.65	8.6	36.5
IPO（Azar 等人，2024年）	1.88	80.8	82.9	-2.58	7.8	35.2
KTO（Ethayarajh 等人，2024年）	1.79	82.1	84.1	-2.71	8.1	36.1
Physio-DPO（我们的方法）	**1.28**	**87.5**	**92.8**	**-3.05**	8.2	**33.8**
相较于标准 *DPO* 的提升	*(-29%)*	*(+7.6%)*	*(+11%)*	*(-15%)*	-	多样的

表1：蛋白质生成结果。模型在 ProGen2-XL 上进行微调，并在 30K 个样本上评估。我们报告 sc-RMSD（↓）、可折叠性（pLDDT > 70, ↑）和能量（↓，相对能量单位）。 †：ESM-2 使用吉布斯采样，计算开销大，且与自回归模型的可比性较低。**粗体**/下划线：最佳/次佳。

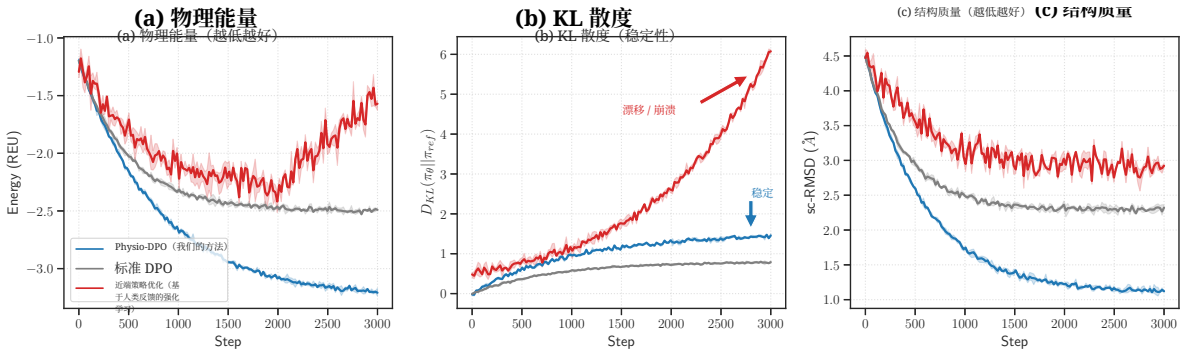

图3：训练动态曲线，(a) 物理能量，(b) KL 散度，(c) sc-RMSD。

5.4 消融研究

表2 展示了 Physio-DPO 的消融实验结果。将困难负样本挖掘替换为随机负样本会导致最大的性能下降，sc-RMSD 从 1.28 埃上升至 2.12 埃，这表明随机负样本提供的梯度较弱且缺乏信息。移除物理信息加权项（退化为标准 DPO）后，可折叠性进一步降低 9.2%，证实对偏好对一视同仁无法反映物理违规的严重程度。最后，将 Sigmoid 加权替换为线性方案会导致性能下降（1.45 埃），这表明无界线性缩放对极端能量差过于敏感，而Sigmoid函数能够实现更稳定、有效的梯度调节。

消融	变化	sc-RMSD ↓	可折叠性 ↑
Physio-DPO	*Sigmoid*函数 +困难负样本	**1.28**	**92.8%**
不含加权	标准 DPO	1.82	83.6%
不含困难负样本	随机负样本	2.12	76.5%
带 线性加权	$\Psi(\delta_E) \propto \delta_E$	1.45	89.1%

表 2：消融研究结果。不含：不包含。

5.5 缓解幻觉

PLM 的一个关键失效模式是生成幻觉——模型对这些序列充满信心（困惑度低），但它们在生物物理上无效。我们在能量与置信度平面中展示了生成序列的分布（图4）。正如预期，标准 DPO 的样本聚集在“高置信度、高能”象限中，这些就是幻觉。Physio-DPO 成功清除了这一象限。样本向“高置信度、低能量”象限移动，表明 Physio-DPO 有效地使模型的置信度与物理现实相一致。

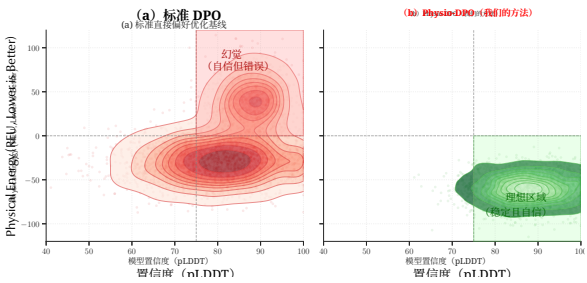

图 4：能量与置信度（pLDDT）平面。

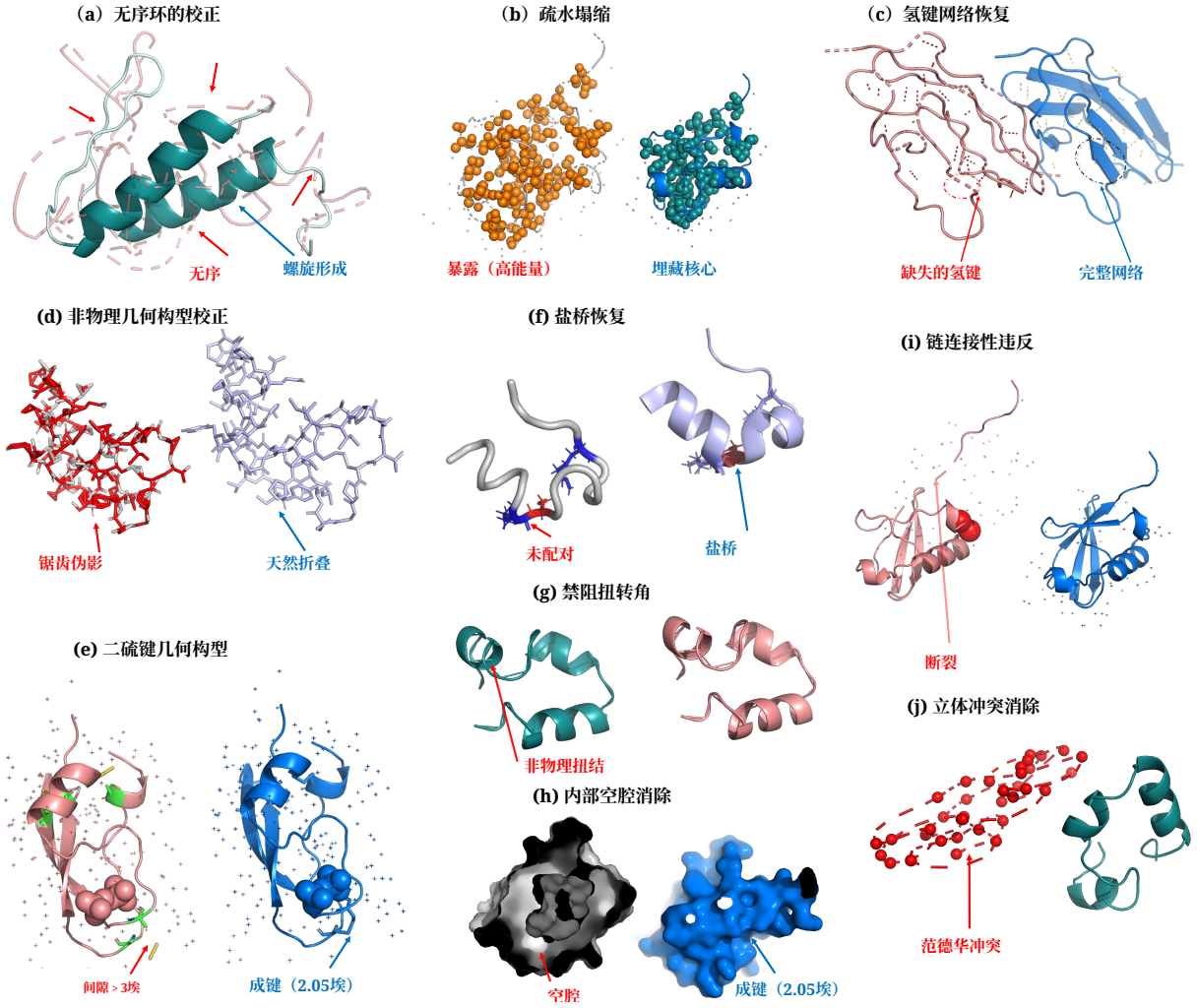

图5：生物物理有效性的综合定性分析。我们比较由 SFT基线（红色/粉色/灰色，左侧）和 Physio-DPO（蓝色/青绿色，右侧）生成的结构。（**a**）Physio-DPO 将无序环区压缩为稳定的螺旋。（**b**）SFT 中暴露的疏水性残基（橙色）经 Physio-DPO 作用后埋入紧密核心（青绿色）中。（**c、e、f**）恢复关键原子相互作用：折叠中的氢键网络、精确的二硫键几何构型（2.05埃），以及静电盐桥。（**d、g、i**）纠正严重的几何违规，包括非物理的“锯齿状”主干网络、禁止出现的扭转角扭曲，以及链连接断裂。（**h、j**）通过消除会降低稳定性的内部空洞和空间位阻冲突（红色），优化堆积密度。

5.6 结构修正分析

为考察 Physio-DPO 如何提升生物物理有效性，图5 对多种结构失效模式进行了可视化比较。尽管 SFT 基线通常能够保留整体拓扑结构，却经常违反细粒度的物理约束。

二级与三级结构稳定性。 Physio-DPO 通过将无序环区压实为结构良好的螺旋（图 5a），并促进疏水核心形成（图 5b），持续提升构象稳定性，从而降低溶剂化能。它还消除了基线结构中观察到的内部空洞（图 5h），使堆积密度更接近天然蛋白质。

原子相互作用恢复。 Physio-DPO 恢复了基线生成结果中经常遭到破坏的关键原子相互作用。这包括恢复 β折叠区域中的氢键网络（图 5c），确保二硫键几何构型正确（图 5e），以及使带相反电荷的残基配对形成稳定盐桥（图 5f）。

几何违规的纠正。能量加权目标还抑制了严重的立体化学违规。与基线相比，Physio-DPO 能纠正非物理的主链扭曲（图 5d）、不允许的扭转角构型（图5g）以及链连接断裂（图5i）。此外，空间位阻冲突也大幅减少（图5j），确保生成的结构符合范德华约束。

方法 （指标：*Spearman's* ρ）	GFP （稳定性）	GB1 （结合）	AAV2 （病毒）	TEM-1 *(耐药性)*	P53 *(抑制因子)*	Avg.
先前基线模型						
ESM-2 (3B) (Lin 等人，2023)	0.62	0.55	0.70	0.63	0.51	0.60
Tranception(Notin 等人，2022)	0.65	**0.68**	0.72	0.69	0.48	0.64
ProGen2-XL (6.4B) 主干网络						
ProGen2-XL（基础版）	0.64	0.56	0.71	0.65	0.53	0.62
SFT	0.66	0.55	0.70	0.67	0.56	0.63
近端策略优化（基于人类反馈的强化学习）	0.62	0.53	0.65	0.61	0.54	0.59
直接偏好优化（标准版）	0.71	0.59	0.73	0.70	0.61	0.67
Physio-DPO（我们的方法）	**0.78**	0.63	**0.75**	**0.76**	**0.70**	**0.72**

表3：ProteinGym 上的零样本适应度预测。报告模型对数似然与实验适应度之间的斯皮尔曼相关系数（ρ）。所有方法均使用 ProGen2-XL。**粗体**：最佳结果；下划线：次佳结果。

5.7 零样本泛化

为评估模型在合成分布之外的泛化能力，我们使用ProteinGym（Notin 等人，2023）在五项具有代表性的检测实验中评估零样本性能（表3），指标为对数似然 π_θ。Physio-DPO 的平均斯皮尔曼相关系数最高，表明物理对齐提升了功能预测能力。在稳定性驱动任务（绿色荧光蛋白和 P53）中，提升最为显著，这与模型有效编码热力学约束的结果一致。相比之下，检索增强基线在 GB1（一项抗体结合任务）上的表现仍然更优，这反映了我们的物理预言机侧重于单体。值得注意的是，近端策略优化会降低零样本性能，而 Physio-DPO 则保留了预训练语义结构。

5.8 超参数敏感性

我们评估 Physio-DPO 对两个关键超参数的稳健性：KL 惩罚系数（β）和物理加权尺度（μ）。

对 KL 惩罚的稳健性。图6（a）比较了 Physio-DPO 和 DPO 在$\beta \in [0.01, 1.0]$下的 sc-RMSD。即使在较小的β值（$\beta = 0.01$）下，Physio-DPO 仍保持稳定，并持续实现较低的 sc-RMSD，这表明通过 Ψ（δ_E）提供的密集物理信息监督信号能够有效规范训练并缓解灾难性遗忘。

物理加权尺度的影响。该尺度 μ 决定能量依赖调制的强度。如图 6（b）所示，μ （< 10）取值较小时，会表现出具有更高能量的类似 SFT 的行为；而取值过大时（> 100），则会过度强调物理能量，导致语言建模性能下降，这体现为困惑度上升。我们确定了一个较宽的最佳范围$\mu \in [20, 50]$，在该范围内 Physio-DPO 在保持出色生成质量的同时实现了较低的物理能量。

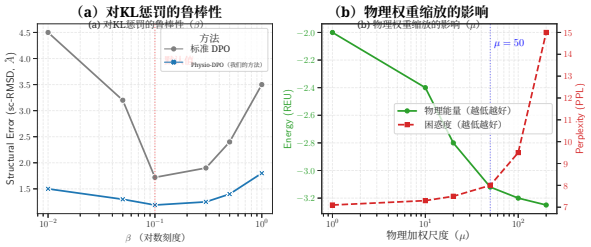

图6：超参数敏感性分析结果。

5.9 补充分析

我们进一步分析了 Physio-DPO 的缩放行为、立体化学有效性和长度鲁棒性。结果表明，随着模型规模增大，其性能有所提升，能够恢复有效的 Ramachandran分布，并在长序列上保持稳定的结构质量。

6 结论

总之，我们提出了 Physio-DPO，这是一种基于物理学的偏好优化框架，旨在使大型蛋白质语言模型与热力学稳定性对齐。我们证明，在生物物理情境下，离散偏好建模并不足够，因为它忽视了能量景观的连续结构。通过将能量大小直接纳入对齐目标，Physio-DPO 引导优化朝着具有物理意义的区分方向进行。我们的结果表明，在对齐阶段融入物理原理，可使大型蛋白质语言模型内化细粒度的生物物理约束，同时不牺牲生成能力或语言能力。

局限性

尽管 Physio-DPO 能有效使蛋白质语言模型与热力学稳定性对齐，但我们当前的研究主要聚焦于单体折叠能量这一物理信号。因此，涉及多态平衡或分子间相互作用的性质并未得到显式优化。此外，我们依赖快速的基于物理学的预言器来近似真实的生物物理能量学，这是蛋白质设计中常见且实用的权衡。值得注意的是，Physio-DPO 不依赖于特定的能量模型，并且可以轻松纳入未来可用的更丰富或针对特定任务的物理信号。

参考文献

Rebecca F Alford、Andrew Leaver-Fay、Jeliazko R Jeli- azkov、Matthew J O'Meara、Frank P DiMaio、Hahn- beom Park、Maxim V Shapovalov、P Douglas Ren- frew、Vikram K Mulligan、Kalli Kappel 等 1 人。2017年。用于大分子建模与设计的 Rosetta 全原子能量函数。《化学理论与计算杂志》，13(6):3031–3048。

Christof Angermueller、David Dohan、David Belanger、Ramya Deshpande、Kevin Murphy 和 Lucy Col- well。2019年。用于生物序列设计的基于模型的强化学习。发表于国际学习表征会议。

Ivan Anishchenko、Samuel J Pellock、Tamuka M Chidyausiku、Theresa A Ramelot、Sergey Ovchin- nikov、Jinghan Hao、Vikas Bafna、Christoffer Norn、Alex Kang、Asim K Bera 等 1 人。2021年。通过深度网络幻觉进行从头蛋白质设计。*Nature*，600(7889):547–552。

Mohammad Gheshlaghi Azar、Zhaohan Daniel Guo、Bi- lal Piot、Remi Munos、Mark Rowland、Michal Valko 和 Daniele Calandriello。2024年。理解从人类偏好中学习的一般理论范式。载于国际人工智能与统计学会议，第 4447–4455 页。PMLR。

Amitava Das、Suranjana Trivedy、Danush Khanna、Yaswanth Narsupalli、Basab Ghosh、Rajarshi Roy、Gurpreet Singh、Vinija Jain、Vasu Sharma、Aish- warya Naresh Reganti 等。2025年。DPO 核函数：一种语义感知、经核函数增强且具有丰富散度的直接偏好优化范式。载于计算语言学协会会议论文集：*ACL 2025*，第 22174–22270 页。

Ahmed Elnaggar、Michael Heinzinger、Christian Dal- lago、Ghalia Rabaan、Florian Burkhardt 和 Burkhard Rost。2021年。Prottrans：通过自监督学习理解生命语言。《IEEE模式分析与机器智能汇刊》，44(10):7112–7127。

Kawin Ethayarajh、Winnie Xu、Niklas Muennighoff、Dan Jurafsky 和 Douwe Kiela。2024年。KTO：将模型对齐视为前景理论优化。*arXiv*预印本 *arXiv:2402.01306*。

Ziqi Fang、Hongbiao Ran、YongHan Zhang、Chensong Chen、Ping Lin、Xiang Zhang 和 Min Wu。2025年。AlphaFold 3：为基础研究和药物开发带来前所未有的机遇。*PrecisionClinical Medicine*，8(3):pbaf015。

Noelia Ferruz、Steffen Schmidt 和 BirteHöcker。2022。ProtGPT2 是一种用于蛋白质设计的深度无监督语言模型。*Nature Communications*，13(1):4348。

Shreya Gopalan 和 Sundar Narayanan。2025年。AlphaFold 3 在本征无序蛋白质中的幻觉：生物过程残基存在无序。*arXivpreprint arXiv:2510.15939*。

Thomas Hayes、Roshan Rao 和其他 1 位作者。2024年。利用语言模型模拟 5 亿年的进化。*bioRxiv*。

Jiwoo Hong、Noah Lee 和 James Thorne。2024年。ORPO：无需参考模型的单体偏好优化。*arXiv* 预印本 *arXiv:2403.07691*。

Edward J Hu、Yelong Shen、Phillip Wallis、Zeyuan Allen-Zhu、Yuanzhi Li、Shean Wang、Lu Wang、Weizhu Chen 和另外 1 位作者。2022。LoRA：大语言模型的低秩适配。*ICLR*，1(2):3。

John B Ingraham、Max Baranov、Zak Costello、Karl W Barber、Wujie Wang、Ahmed Ismail、Vincent Frappier、Dana M Lord、Christopher Ng-Thow-Hing、Erik R Van Vlack 和另外 1 位作者。2023。利用可编程生成模型照亮蛋白质空间。*Nature*，623(7989):1070–1078。

Ronald M Levy、Allan Haldane 和 William F Flynn。2017年。蛋白质进化、结构与功能的 Potts模型。*Current Opinion in Structural Biology*，43:55–62。

Zeming Lin、Halil Akin、Roshan Rao、Brian Hie、Zhongkai Zhu、Wenting Lu、Nikita Smetanin、Robert Verkuil、Oana Kabeli、Yaniv Shmueli 等 1 位作者。2023年。利用语言模型进行进化尺度的原子级蛋白质结构预测。*Science*，379(6637):1123–1130。

Sidney Lyayuga Lisanza、Jacob Merle Gershon、Samuel WK Tipps、Jeremiah Nelson Sims、Lucas Arnoldt、Samuel J Hendel、Miriam K Simma、Ge Liu、Muna Yase、Hongwei Wu 等 1 位作者。2025年。利用 RosettaFold 序列空间扩散进行多状态功能性蛋白质设计。*Nature Biotechnology*，43(8):1288–1298。

Shunyu Liu、Wenkai Fang、Zetian Hu、Junjie Zhang、Yang Zhou、Kongcheng Zhang、Rongcheng Tu、Ting-En Lin、FeiHuang、Mingli Song 等人。2025年。直接偏好优化综述。*arXiv* 预印本 *arXiv:2503.11701*。

Ali Madani、Ben Krause、Eric R Greene、S Subramadian 等人。2023。大语言模型能够生成涵盖多种不同家族的功能性蛋白质序列。*NatureBiotechnology*，41(8):1099–1106。

Zhiyu Mei、Wei Fu、Kaiwei Li、Guangju Wang、Huanchen Zhang 和 Yi Wu。2025年。Real：通过参数重新分配实现大语言模型的高效基于人类反馈的强化学习训练。《机器学习与系统会议论文集》，7。

Yu Meng、Mengzhou Xia 和 Danqi Chen。2024年。Simpo：采用无参考奖励的简单偏好优化。神经信息处理系统进展，37：124198–124235。

Erik Nijkamp、Jeffrey A Ruffolo、Eli N Weinstein、Nikhil Naik 和 Ali Madani。2023。Progen2：探索蛋白质语言模型的边界。《细胞系统》，14(11)：968–978。

Pascal Notin、Mafalda Dias、Jonathan Frazer、Javier Marchena-Hurtado、Aidan N Gomez、Debora Marks 和 Yarin Gal。2022。Tranception：利用自回归变换器和推理时检索预测蛋白质适应度。载于国际机器学习会议，第16990–17017页。PMLR。

Pascal Notin、Aaron Kollasch、Daniel Ritter、Lood Van Niekerk、Steffanie Paul、Han Spinner、Nathan Rollins、Ada Shaw、Rose Orenbuch、Ruben Weitz-man 和其他 1 位作者。2023。ProteinGym：用于蛋白质适应度预测和设计的大规模基准测试。《神经信息处理系统进展》，36:64331–64379。

Rafael Rafailov、Archit Sharma、Eric Mitchell、Christo- pher D Manning、Stefano Ermon 和 Chelsea Finn。2023。直接偏好优化：你的语言模型暗中是一个奖励模型。《神经信息处理系统进展》，36:53728–53741。

Alexander Rives、Joshua Meier、Tom Sercu、Siddharth Goyal、Zeming Lin、Jason Liu、Demi Guo、Myle Ott、C Lawrence Zitnick、Jerry Ma 和其他 1 位作者。2021年。通过将无监督学习扩展至 2.5 亿条蛋白质序列，生物结构和功能得以显现。《美国国家科学院院刊》，118(15):e2016239118。

John Schulman、Filip Wolski、Prafulla Dhariwal、Alec Radford 和 Oleg Klimov。2017年。近端策略优化算法。*arXiv* 预印本 *arXiv:1707.06347*。

Ashish Vaswani、Noam Shazeer、Niki Parmar、Jakob Uszkoreit、Llion Jones、Aidan N Gomez、Łukasz Kaiser 和 Illia Polosukhin。2017年。注意力就是一切。《神经信息处理系统进展》，第 30 卷。

Joseph L Watson、David Juergens、Nathaniel R Bennett、Brian L Trippe、Jason Yim、Helen E Eisenach、Woody Ahern、Andrew J Borst、Robert J Ragotte、Lukas F Milles 等另外 1 人。2023。从头设计蛋白质结构与功能的 RFdiffusion。*Nature*，620(7976):1089–1100。

Jiayi Zhou、Jiaming Ji、Josef Dai 和 Yaodong Yang。2025年。序列到序列奖励建模：通过语言反馈改进基于人类反馈的强化学习。载于*AAAI*人工智能会议论文集，第 39 卷，第 27765–27773 页。